THE

ONTARIO WATER RESOURCES

COMMISSION

A Statistical Approach to
Water Quality
Monitoring
Programs

1969

Statistical Approach

To

Water Quality

Monitoring Programs

by

Dr. Syed S. Ahmed Rizvi

Water Quality Surveys Branch

1969

Ontario Water Resources Commission

ABSTRACT

Water quality data is very important in a water management program. The data is used to detect changes in water quality, to ascertain the amount of chemicals flowing into lakes, to give detailed information on the cause and effect relationships of the factors affecting water quality. Such survey programs are expensive to operate so attempts must be made to eliminate the necessity of taking some samples which are producing similar water quality information and still have data adequately describing water quality variations.

In this report, spatial and temporal factors affecting water quality are discussed. Analytical statistical techniques, both parametric and non-parametric, are presented to determine whether there are significant differences between data of consecutive months. The selection of parameters and groups is discussed and the techniques are applied to data from Duffin Creek. The results are presented and various recommendations are made.

TABLE OF CONTENTS

INTRODUCTION

With the growing demand for good water, it is necessary to determine the quality of existing water supplies in the different areas covered in a water management program and to assess the changes, if any, taking place in the water quality. The efficiency and adequacy of the monitoring program should be given particular attention in order to ensure a sufficiently large supply of data and to indicate what measurements should be made in future monitoring programs. To carry out the monitoring program economically, with existing survey techniques, it is necessary to determine how often and how many samples should be taken for the analysis and interpretation of data.

The Water Quality Surveys Branch is conducting a regular water quality monitoring survey program of the river basins in the Province of Ontario. The purpose of the program is to ascertain the amount of various chemicals flowing into the Great Lakes, and to develop, based on the survey data, a program in the watersheds for maintaining clean water. Collected over several years, the water quality data can be used to indicate long-term environmental changes in watersheds and to give detailed information on the cause and effect relationships of the factors affecting water quality. Effective water management rests on a sound water quality monitoring system. Since

such systems are expensive to operate, it is important that the sampling program is designed to produce sets of data which can be interpreted using reliable statistical techniques and is the most efficient one possible.

The question often asked is how many observations should be made to ensure that the risk of making a wrong decision is acceptably small. Some risk is, of course, always present because of the random variation in results, but the risk becomes smaller and smaller as the number of observations increases. It is clear, however, that for reasons of economy of time and effort, the extent of the monitoring program should be kept to a minimum, consistent with the maximum risk we are prepared to accept of making a wrong decision. In practice, there will be a certain optimum rate of sampling. However, it is often better to evaluate the cost of sampling and the economic implications of making a wrong decision before one can consider decision-making on the basis of a specified risk of being wrong. A partial answer to this question is found in the theory of sampling which will be discussed here briefly.

Before commencing on the adequacy of background data or on data which should be developed in the future, it is necessary to evaluate the existing sampling data. This evaluation can help to determine whether the data of particular months and years is or is not contributing

significantly to the total water quality information. It can also indicate whether or not it is possible to eliminate the necessity of taking some samples which are producing similar water quality information and still have data which adequately describe water quality variations. Therefore, this study was initiated to test various known statistical techniques for their applicability in detecting similarities and/or differences in the water quality information collected during the five year period (1964-68).

I VARIATION OF WATER QUALITY DATA

Space and time are two elements which affect all important water quality constituents. Therefore, an insight from this point of view may provide some information on the optimal monitoring conditions.

(a) Spatial Variation

The quality of water in a flowing stream varies from place to place within the system, chiefly as a function of the physiographic and geological characteristics of the tributary drainage, the distribution and type of man-generated and natural sources of upstream pollution and the hydrologic properties of the basin. In general, it can be said that the quality of stream water is best at the highest point of the drainage, where it is closest in time and space to its original form as precipitation, and has had minimum exposure to man. As the water moves downhill and the tributary drainage area is expanded, the potential for pollution increases both through natural causes and human activity, since the number and variety of pollutants would be expected to increase. With the introduction of pollutants of domestic and industrial origin, usually centered at points within the system rather than dispersed as in the case of natural pollution input, the stream experiences abrupt changes in its quality. The greatest rate of change occurs close to

pollution sources. These changes may be attenuated somewhat

because of mixing with greater quantities of dilution water from un-

polluted sources as the drainage area is expanded. Generally, it

can be assumed that spatial variation in the quality of streams is

one-dimensional in most instances. Changes along the vertical

or transverse axes are relatively minor in comparison with changes

along the axis of flow.

(b)   Temporal Variation

Temporal variations in stream quality are due to a large

number of time-dependent factors, most of which are related either

to the annual hydrologic cycle or to the various monthly, weekly or

diurnal cycles possible. Some of these factors, such as precipitation,

are predominantly random in character and consequently exert a more

dramatic influence on quality than will be observed with strictly

periodic factors. Because of the random nature of these factors,

related quality responses may be most appropriately evaluated in

statistical form.

It should be noted that hydrologic factors and quality res-

ponses associated with them are composites of random and periodic

components, the degree of randomness varying with time and space

scales. As the scales increase, random effects become less important

and the responses to periodic factors tend to be more significant.

Thus, it can be appreciated that the suitability of statistical techniques in a water quality monitoring program in rivers and streams is not simply a matter of arbitrary choice. Careful consideration must be given to the size of the sample, the time interval and the nature of hydrologic phenomena which either directly or indirectly affect quality. One can say that in the headwaters of a stream, where the degree of natural or artificial pollution is likely to be minimal, the diurnal cycle may not be as important as the annual cycle. Low in the drainage area in a biologically more active environment diurnal patterns such as the daily dissolved oxygen cycle associated with photosynthesis may become controlling. Annual quality fluctuations, closely tied in with hydrologic phenomena, tend to be attenuated lower in the drainage and increased in amplitude high in the drainage.

II ANALYTICAL TECHNIQUES

The statistical technique selected for this case study, known as analysis of variance, is probably the most powerful procedure in the field of experimental statistics. The subject of the analysis of experimental data is so vast that one can at best touch upon some of the basic ideas here.

(a) Parametric Techniques

In general, the variation in sample values consists of two types of fluctuations: random (or sampling) fluctuations and systematic fluctuations. The function of analysis of variance is to find the systematic variations between means of different samples and to determine whether they are real or, in statistical language, whether they are significant. Another purpose is to determine how specific sources of variation contribute to the total variance of a quantity, and to test whether the effect of a particular factor is real or likely to have arisen as a result of random errors. In the simplest form of this analysis, the mean of two samples can be tested for a significant difference; that is, whether or not both are from the same population with a mean u. This is done using the t-test which can be found in any book on statistics.[1]

---

[1] Freund, J.E., Mathematical Statistics, Prentice Hall, 1962. p.266-269.

Although the basic computations for the analysis of variance are usually quite straight-forward and described adequately in most texts, the interpretation of the computations may become difficult for more complex situations since it depends on the purpose of the analysis and the assumptions made in setting up the experimental model. The underlying assumption for these techniques is that the values within a column are normally distributed. Generally, in the analysis of variance, the null hypothesis is formulated. This requires a test to ascertain whether or not the means of the n samples (or columns) within each group are similar and represent the same population. The method used to test this hypothesis is based on the analysis of the total variability of the data. If the null hypothesis is true, then

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2$$

which is (nk-1) times the variance of all the data, is due entirely to chance variation. Here $\bar{x}$ is the mean of all the data; namely,

$$\bar{x} = \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}$$

If the null hypothesis is not true, then part of this sum of squares can be attributed to differences among the means of the n populations. To be more specific one can write:

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + n \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2$$

(one-way analysis of variance)

or $$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2 = n \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2 + k \sum_{j=1}^{n} (\bar{x}_j - \bar{x})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i - \bar{x}_j - \bar{x})^2$$

(two-way analysis of variance)

It is customary to refer to the above equations as:

$$SST = SSE + SSB$$

$$SST = SSA + SSB + SSE$$

where SST is the total sum of squares, SSE is error sum of squares, SSB as sum of squares between sample B's, and SSA as sum of squares between A's. From these equations one sees that SST, which is a measure of the total variability of the data, has been broken down into the sum of two or three components. SSE measures chance variation (the variability within the samples or experimental error) regardless of whether the null hypothesis is true. SSB measures chance variation if the null hypothesis is true, but it can be attributed in part to differences among the population means if the null hypothesis is false. In the two-way analysis equation, the total variability of the data is broken down into terms attributed to differences among the A's (rows $i = 1, 2...k$) and B's (columns $J = 1,2...n$) and the chance error. In the two-way analysis, there are two null hypotheses based on the A's and B's; that is, there is no significant difference among the A's and there is no significant difference among the B's. The null hypothesis concerning A is based on the statistic:

$$F_A = \frac{SSA/\text{degrees of freedom between A's}}{SSE/\text{degrees of freedom between B's}} = \frac{SSA/(k-1)}{SSE/(k-1)(n-1)} = \frac{(n-1)SSA}{SSE}$$

and concerning B is based on:

$$F_B = \frac{SSB/(n-1)}{SSE/(n-1)(k-1)} = \frac{(k-1)\,SSB}{SSE}$$

We reject the null hypothesis if F_A or F_B are greater than F_{α} , $(k-1),(n-1)(k-1)$ and F_{α} , $(n-1), (n-1)(k-1)$, (tabulated) respectively. In the above model we did not allow for the possibility that there might be joint effects of the two variables; that is, so-called interactions. The above values of F_A and F_B may also be written as:

$$F_A = \frac{MSA}{MSE} \geq F \text{ tabulated} \qquad \text{(rows)} \qquad \text{rejection of null hypothesis}$$

$$F_B = \frac{MSB}{MSE} \geq F_{tabulated} \qquad \text{(columns)} \qquad \text{rejection of null hypothesis}$$

where MSA, MSB, MSE are the mean squares for A (row effects), for B (column effects) and residual sources respectively. If the hypothesis that there are no row effects or no column effects or both is rejected, some other statistical method[1] can be applied to get a confidence interval on the difference between two row effects or two column effects.

Analytical techniques for comparing the means within a group and the variance (measure of the variation from the mean) are well-established and documented. The techniques used in the case study are statistically strong, if an adequate number of samples are available within a group. Analysis of variance for multiple factors is really only an extension of the two-way analysis.

---

[1] Bowker, A. H. and Lieberman, G. J., Engineering Statistics, Prentice Hall, New York. p.323 and 295.

(b) Non-parametric Techniques

The parametric technique for testing whether several samples have come from identical populations is the analysis of variance or F test. The assumption for the statistical model underlying the F test are: that the observations are independently drawn from normally distributed populations; that the populations all have the same variance; and that the means in the normally distributed populations are linear combinations of "effects" due to rows and columns; that is, that the effects are additive. However, if normality cannot be assumed and if the other assumptions are unrealistic for the data available, it is necessary to investigate the non-parametric techniques of analysis. This also would increase the generality of the findings by avoiding the assumptions made above.

Although there is no underlying assumption of normality in non-parametric methods, a requirement for a sufficiently large number of values exists in order to obtain results comparable to the parametric methods. One must also keep in mind that if parametric methods are applied to non-normal distributions, some invalid conclusions may result. Both the Mann-Whitney U and Wald-Wolfowitz runs tests are useful for a one-way analysis of two samples. As mentioned earlier, one-way analysis is applied when samples are independent, that is when there is no row effect on the values in the columns.

If one is interested simply in testing whether two samples come from the same population or whether they represent populations which differ in central tendency (mean), the Mann-Whitney U-test would be a more powerful test than the Wald-Wolfowitz runs test because it is specifically designed to disclose differences of this type. The runs test is designed to disclose differences of any type (central tendency, variability, skewness, etc.) and is thus less powerful in disclosing the particular kind of difference needed here. (This test can be applied if sufficiently large samples are available).

The power efficiency of the Mann-Whitney U-test is 95 per cent while that of the Wald-Wolfowitz runs test is not known. The concept of power efficiency is concerned with the amount of increase in sample size which is necessary to make these tests as powerful as corresponding parametric tests. Thus, one can avoid having to meet some of the assumptions of the more powerful, parametric tests without losing power by simply choosing a different test and drawing a larger total number of samples (N). In other words, by choosing another statistical test with fewer assumptions in its model and thus with greater generality than the T and F tests, and by enlarging N, one can avoid having to make the assumptions mentioned earlier, and still retain an equivalent power to reject the null hypothesis.

The choice of non-parametric techniques is much more limited for the n related or unrelated samples than for two samples. After looking at the data available in each month, it was decided that the Friedman two-way analysis for n related samples (where $n > 2$), and the Kruskal-Wallis one-way analysis were useful for testing the null hypothesis that the n samples have been drawn from the same population.

Both tests are similar in that ranks are used. However, the Kruskal-Wallis test does not consider row effects as the Friedman test does. The tests assume that the variable under study has an underlying continuous distribution. The Kruskal-Wallis method can be used where there is a variable number of samples in each group (or column). Furthermore, this method, based upon comparison of averages, results in a power efficiency of 95.5 per cent compared to the parametric analysis of variance.

III METHOD AND DISCUSSION

The purpose of this study was to see if the number of samples taken each month or year could be reduced, thereby achieving more economical operations. This reduction should be made without reducing the usefulness of this data in the future.

(a) Selection of Parameters

The job of selecting the reliable water quality parameters which are important in the monitoring program is a difficult one. As a general guide, each parameter must be measured frequently and reliably and must give a good indication of pollution in the streams. For these reasons, the following parameters were selected:

Dissolved Oxygen (DO)	Suspended Solids
Coliforms	Nitrates + Nitrites
Total Solids	$(NO_3 + NO_2)$
5-Day Biochemical Oxygen	Total Kjeldahl Nitrogen
Demand (BOD_5)	Total Phosphates

The reliability and accuracy of these selected water quality survey parameters is difficult to determine. A crude measure of the variability of the values is the ratio of the maximum to the minimum value. A low ratio, say, less than 15, would characterize a well-regulated stream with comparatively little expectation of quality variation due to hydrologic influences. Hence a minimal sampling program may be called for. In contrast, a ratio in excess of 100

would typify a stream with large variations of a random nature might
be expected. Such a stream could require sampling as often as once
a week in order to adequately define the random variables in quality
over the annual cycle. The range and ratio are shown in Table 1 and
Table 2A.

The most difficult indicator to work with is the coliform count,
since it normally varies by orders of magnitude. Averages are depen-
dent almost entirely on a few high values and the corresponding ratios
are misleading. Therefore, analysis of this data is a little more
difficult. The other difficulty involves parameters where the minimum
value equals zero. The ratio equals infinity and this is unrealistic.

The accuracy of the measurements is affected by low concen-
trations (as in the case of $NO_3 + NO_2$ and total phosphate) and the
limitations of the standard measurement techniques used. The tech-
niques for all the parameters except total phosphate remained un-
changed so that the reliability and accuracy did not change. Methods
of measurement should be consistent so that any possible type of
error remains constant. Whenever there is an improvement in tech-
nique, further complications are introduced into the analytical studies.
Though general improvements were made in the phosphate determination,
it was assumed the accuracy of determining this parameter remained
the same.

For this case study, the water quality data collected at Duffin Creek was chosen because of the long records available. This stream was representative of both a rural and an urban (industrial) drainage basin. The particular station chosen was near the mouth of the stream.

(b) Selection of Groups

After some thought it was decided that the data for each selected parameter could be divided into groups of months. Thus, months within each group would be compared with one another to see if the data for each month differed significantly. Each group contained n months where n could vary from 2 ... 12, depending upon the grouping. As the monitoring data was available from 1964-1969, each month had at least 4 or 5 samples. The final grouping of months is shown in Figure 1. Each parameter usually had a slightly different grouping pattern though general trends were clearly visible.

The initial selection of the groups was somewhat intuitive and based on trial and error. To begin the selection, visual aids were used; that is, plotting the monthly values and moving averages for two or three months. The change in the slope of the line indicated some grouping possibilities which were used as a starting point.

In checking and arranging the monthly data for the two-way analysis of variance, it was noticed that in the period, 1964-1969,

some readings were missing; in other cases, there were two or three

readings per month (it was decided to use only one observation per

month). Naturally, two questions arose. What values should be

substituted for the missing values and which value of two or more

monthly values should be taken; for example, the highest or the

lowest? For the first question, it was decided to substitute the

mean value for the missing values, so that the substituted values

did not greatly affect the comparison of group means, especially in

the case of the two-way analysis of variance since the two-way

classification required an equal number of readings in each column.

As for the second question, it was decided to use the highest of the

readings in a particular month because high values were more

critical from a pollution point of view in this study. The one excep-

tion was DO where low values are more critical in pollution deter-

minations. Therefore, low values were used in the two-way analysis

of variance.

(c) Application of Parametric Analysis of Variance

After the groups were decided and the values selected, a two-

way analysis of variance was applied to the data. The two-way

analysis of variance indicates the discrepancies among the means

of different months and years within each group; that is, whether

they may reasonably be attributed to chance or indicative of differences

among the means of the corresponding populations. Alternatively, one can ask whether it is possible to test if two or more sample means (means of each month) are likely to have come from the same hypothetical population with mean u and variance   .

In this case study, the months were used as columns and the years as rows and one reading per combination was available. Tables 1A, 3A, and 3B have been prepared in such a way that either of the two effects, column effect or row effect, can be studied by two-way analysis. The important thing is the degree of freedom shown. Tables 1 and 2A also show some further computations including a one-way analysis done on the same data but excluding the substituted mean values for missing ones using an unequal sample method.[1] This information along with the applied F test was used to show if there was any evidence to support the hypothesis that the means of the months within each group were significantly different. A significant difference was indicated by SD; no significant difference by NSD in all the tables.

Further, a one-way analysis of variance was also performed on all the data available, and this required no substitution of values since the columns did not have to have the same number of readings (i.e. using the unequal sample method). [1] The results are shown in

---

[1] Introduction to Probability and Statistics, Alder and Roessler, Freeman 1969 (Fourth Ed.) p. 253.

Table 2. It should be noted that a level of significance, α , of 0.05 was used for all the tests applied, both parametric and non-parametric.

In view of the lack of information on the population distribution and the unavailability of proven transforms, it was felt that no transformation should be applied. Nevertheless, it should be borne in mind that if the transformations were known, they could be applied to small samples (only 5-10 values) with confidence.

One further check was made in this study, though the actual results will not be reported here. Instead of using the highest values, the lowest of the values for each month were substituted into the final grouping for all the parameters except DO. A full analysis was done on these data as above for the high values and the results were compared. It was found that the results were more ambiguous and misleading than when high values were used though in most cases there was good agreement. This was particularly true where values were very low (especially nitrate and nitrite $(NO_2 + NO_3)$)in May, June, July and August where there were a large number of zero values and values of the same magnitude). This confirmed the original decision to use high values for the two-way analysis.

If a group of three or more months (columns) showed no significant differences among the columns, any combination of the months

within the group also showed no significant difference for both new groups.

(d) Application of Non-parametric Methods

Generally, parametric tests are stronger than non-parametric tests, especially for given limited values. One needs more values to make the non-parametric tests as powerful as the parametric tests. However, to test results based on some of the assumptions implied in the use of the parametric tests, the non-parametric tests mentioned above were used as a check.

In applying the non-parametric techniques to the groups, several factors came into consideration. Two-way non-parametric tests; for example, the Friedman two-way analysis of variance by ranks, could only be applied in cases of three or more columns per group. There was insufficient data for a two-way analysis of groups with only two columns, so that tests like the Wilcoxon matched pairs test could not be used. For this reason, it was discovered that the one-way tests were more useful to this study, especially since all the available data could then be utilized.

There were three one-way non-parametric tests used: the Kruskal-Wallis one-way analysis of variance by ranks, the Mann-Whitney U-test, and the Wald-Wolfowitz runs test. The Kruskal-Wallis test could be used only for three or more columns per group,

but gave results agreeing well with the corresponding parametric tests. The Mann-Whitney U-test was applied to groups with only two columns and the runs test was applied wherever there was sufficient data, as a check on the Mann-Whitney test. Generally, the Mann-Whitney U-test was regarded as the stronger test.

In using the non-parametric tests it was important to ensure that the tests were two-tailed. In one case, for the Mann-Whitney U-test, the table of probabilities (Table J, p.271-273[2]) given in the text was only for a one-tailed test. The appropriate values had to be doubled for a two-tailed test. In using any tables it is important to check whether they are two-tailed or one-tailed.

The main difficulty encountered in the use of the one-way two sample tests was the occurrence of ties; that is, two or more readings of the same magnitude. There was no provision for either deciding a tie or considering it in the calculations. This presented some difficulty in making a decision on whether a significant difference existed. It seemed best to consider several possible arrangements but not for all. After considering the possibilities it was decided there was actually no significant difference. In general, the results of the non-parametric tests closely paralleled those of the parametric tests. Tables 1A and 2 shows the results of all the tests.

---

[2] Non-parametric Statistics for the Behavioral Sciences, S. Siegel, McGraw-Hill, p.271-273.

IV SUMMARY AND CONCLUSIONS

In this study, the most important statistical technique used was the analysis of variance, both parametric and non-parametric, one-way and two-way. On the basis of this investigation, one-way analysis of variance is more useful than the two-way analysis, largely because all the data can be used. There is no necessity for an equal number of readings in each column, and no need to substitute possibly misleading values for missing data. This also eliminates the need for deciding which one of several readings in a particular month to use. However, two-way analysis is useful in determining the nature of the row effects. There were cases where significant differences occurred between rows, which indicated a significant difference from year to year, though this was rare. These effects are shown in Table 3B. When two-way analysis is used, the highest values per month should be used except in the case DO where lowest values should be used. If a value is missing the mean value should be substituted.

One may be tempted to use non-parametric tests as the main method of analysis because of several seeming advantages including ease of computation and the assumption of no particular frequency distribution. However, if the data is from a well-behaved distribution, we throw away information by using the non-parametric tests

instead of parametric tests. This may mean we can prove a significant difference between two or more groups by a parametric technique but not by a non-parametric technique. However, it is still best to perform some non-parametric tests along with the parametric techniques.

In case studies like this more information, possibly obtained using recording-type meters, is required to determine the population distribution of the variables. Once this is known, the validity of the various forms of testing can be implicitly determined. This kind of population distribution information is required for all the parameters to provide the best possible understanding of the analytical techniques, and may be worthwhile obtaining sometime in the future.

Based on the short-term data used in this study, the following methods are best applied to the analysis of the data:

Parametric: Analysis of Variance   --   one-way, two or more
                                                    unequal samples, using
                                                    all availabe data.
                                          --   two-way, two or more
                                                    equal samples, using
    high values (except in
    case of DO).

Non-parametric: Mann-Whitney U-test --   one-way, two unequal
    samples, using all
    available data.
                  Kruskal-Wallis Analysis of Variance
    --   one-way, three or more
    unequal samples.

In this study, all the tests mentioned above gave generally consistent results, even though the nature of the tests varied to some extent.

Though the grouping of months varied from one parameter to another, definite patterns were discernible. This case study revealed that it was possible to produce a grouping of months which apply in general to all eight parameters:

> February, March, April
> May, June
> July, August
> September, October
> November, December, January.

The above grouping could be applied to other streams and stations to confirm the results of this case study. Further study is under consideration to support the validity of this grouping.

This study showed that it is possible to reduce the number of monthly readings required for the particular station studied thereby achieving some financial gain. At the same time no information which is required for decision-making is lost. Of course, consideration should also be given to social, economic and political factors involved in making the final decision on the nature of the monitoring program. Based on the statistical analysis of this short-term data, it is possible to reduce the number of samples required. Of course one must remember that any investigation based on short-term data is not as valid as that based on long-term data.

The analysis presented here is based on temporal readings but it can be applied to spatial values as well; that is, finding significant differences between readings at different places, say, on a lake. As more information on the population distribution becomes available, different testing methods could be applied which may change the grouping pattern. However, at present, it is possible to apply the above techniques to other streams and stations with some mathematical confidence.

APPENDIX. *NOTATION

α = level of significance, probability of a Type I error;

BOD_5 = 5-day, $20^{\circ}C$, Biochemical Oxygen Demand;

DO = Dissolved Oxygen

F = the F test: parametric analysis of variance;

F_A = $F_{\alpha', (k-1), (n-1)(k-1)}$ = $\dfrac{MSA}{MSE}$ = $\dfrac{(m-1)\,SSA}{SSE}$ (rows);

F_B = $F_{\alpha', (n-1), (k-1)(n-1)}$ = $\dfrac{MSB}{MSE}$ = $\dfrac{(k-1)SSB}{SSE}$ (columns);

k = the number of rows;

MSA = mean sum of squares (rows) = $\dfrac{SSA}{k-1}$;

MSB = mean sum of squares (columns) = $\dfrac{SSB}{n-1}$;

MSE = mean sum of squares (residual or error) = $\dfrac{SSE}{(n-1)(k-1)}$;

μ = the population mean;

n = the number of samples (columns);

N = the total number of random variables;

$NO_3 + NO_2$ = nitrate plus nitrite;

NSD = no significant difference;

ppm = parts per million;

SD = significant difference;

SSA = sum of squares (rows);

SSB	=	sum of squares (columns);
SSE	=	sum of squares (residual or error);
SST	=	sum of squares (total);
T	=	Student's t test: a parametric test;
x, x_{ij}	=	the random variables;
$\bar{x}$	=	the sample mean.

BIBLIOGRAPHY

1. Freund, J. E., Mathematical Statistics, Prentice-Hall, 1962, New York.

2. Bowker, A. H. and G. J. Lieberman, Engineering Statistics, Prentice Hall, New York.

3. Siegel, S., Non-parametric Statistics for the Behavioral Sciences, McGraw-Hill, New York.

TABLE I

STATISTICAL RESULTS WITH GROUPING
FOR ALL VALUES
DUFFIN CREEK

WATER QUALITY PARAMETERS	GROUPS	RANGE	RATIO: HIGH/LOW	PARAMETRIC 1 - WAY	KRUSKAL- WALLIS	MANN - WHITNEY	WALD - WOLFOWITZ
DISSOLVED OXYGEN	FEB., MAR., APR.	5.0, 14.0	2.80	N S D	N S D	*	*
	MAY, JUN.	8.0, 13.0	1.62	N S D	*	N S D	N S D
	JUL., AUG., SEPT.	7.0, 12.4	1.77	N S D	N S D	*	*
	OCT., NOV.,	4.0, 13.0	3.25	N S D	*	N S D	S D
	DEC., JAN.	6.0, 16.0	2.67	N S D	*	N S D	N S D
COLIFORMS	FEB., MAR., APR.	212, 21000	99.10	N S D	N S D	*	*
	MAY, JUN.	100, 7000	70.00	N S D	*	N S D	N S D
	JUL., AUG.	170, 14000	82.40	N S D	*	N S D	N S D
	SEPT., OCT.	252, 36000	143.00	N S D	*	N S D	N S D
	NOV., DEC., JAN.	280, 8000	28.60	N S D	N S D	*	*
TOTAL SOLIDS	FEB., MAR., APR.,	270, 962	3.56	N S D	N S D	*	*
	MAY, JUN.	234, 400	1.71	N S D	*	N S D	S D
	JUL., AUG.	264, 726	2.75	N S D	*	N S D	N S D
	SEPT., OCT.	222, 398	1.79	N S D	*	N S D	N S D
	NOV., DEC., JAN.	290, 494	1.71	N S D	N S D	*	*
BOD (5 DAY)	FEB. MAR., APR.	0.9, 22.0	24.50	N S D	N S D	*	*
	MAY, JUN., JUL.,AUG.	1.0, 32.0	32.00	N S D	N S D	*	*
	SEPT., OCT.	0.2, 3.2	16.00	N S D	*	N S D	N S D
	NOV., DEC., JAN.	0.4, 3.3	9.00	N S D	N S D	*	*
SUSPENDED SOLIDS	FEB., MAR., APR.	10, 753	75.30	N S D	N S D	*	*
	MAY, JUN.	5, 44	8.8	N S D	*	N S D	N S D
	JUL., AUG.	13, 44	3.39	N S D	*	N S D	N S D
	SEPT., OCT.	5, 50	10.00	S D	*	N S D	N S D
	NOV., DEC., JAN.	5, 130	26.00	N S D	N S D	*	*

TABLE 1 (CONTINUED)

STATISTICAL RESULTS WITH GROUPING

FOR ALL VALUES

DUFFIN CREEK

WATER QUALITY PARAMETERS	GROUPS	RANGE	RATIO: HIGH/LOW	RESULTS PARAMETRIC 1 - WAY	NON - PARAMETRIC KRUSKAL-WALLIS	MANN - WHITNEY	WALD - WOLFOWITZ
NITRATE & NITRITE	FEB., MAR., APR.	0.12, 5.00	41.60	N S D	N S D	*	*
	MAY, JUN., JUL., AUG.	0.00, 0.47	-	N S D	N S D	*	*
	SEPT., OCT.	0.00, 12.50	-	N S D	*	N S D	N S D
	NOV., DEC., JAN.	0.13, 1.28	9.85	N S D	N S D	*	*
KJELDAHL NITROGEN	FEB., MAR., APR.,	0.32, 2.00	6.25	N S D	N S D	*	*
	MAY, JUN., JUL., AUG.	0.00, 1.70	-	N S D	N S D	*	*
	SEPT., OCT.	0.26, 0.78	3.00	N S D	*	N S D	N S D
	NOV., DEC., JAN.	0.30, 0.78	2.60	N S D	N S D	*	*
TOTAL PHOSPHATE	FEB., MAR., APR.	0.08, 2.10	26.20	N S D	N S D	*	*
	MAY, JUN.	0.07, 0.90	12.90	N S D	*	N S D	N S D
	JUL., AUG., SEPT., OCT.	0.04, 19.0	475.00	N S D	N S D	*	*
	NOV., DEC., JAN.	0.02, 0.58	26.00	N S D	N S D	*	*

N S D - NO SIGNIFICANT DIFFERENCE

S D - SIGNIFICANT DIFFERENCE

* - TEST NOT APPLICABLE.

TABLE IA

STATISTICAL RESULTS WITH GROUPING
(USING MAXIMUM VALUES)#
DUFFIN CREEK

WATER QUALITY PARAMETERS	GROUPS	RESULTS						
		PARAMETRIC			NON-PARAMETRIC			
		1-WAY	2 - WAY DIFF.IN COL. B's	2 - WAY DIFF. IN ROWS A's	FRIEDMAN 2- WAY	KRUSKAL - WALLIS 1 - WAY	MANN - WHITNEY	WALD - WOLFOWITZ
DISSOLVED OXYGEN	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN.	N S D	S D	S D	*	*	N S D	N S D
	JUL., AUG., SEPT.	N S D	N S D	N S D	N S D	N S D	*	*
	OCT.,, NOV.	N S D	N S D	N S D	*	*	N S D	S D
	DEC., JAN.	N S D	N S D	N S D	*	*	N S D	N S D
COLIFORMS	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN.	N S D	N S D	N S D	*	*	N S D	N S D
	JUL., AUG.	N S D	N S D	N S D	*	*	N S D	N S D
	SEPT., OCT.	N S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	N S D	N S D	N S D	*	*
TOTAL SOLIDS	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN.	N S D	N S D	N S D	*	*	N S D	S D
	JUL., AUG.	N S D	N S D	N S D	*	*	N S D	N S D
	SEPT., OCT.	N S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	N S D	N S D	N S D	*	*
5-DAY	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN., JUL.,AUG.	N S D	N S D	N S D	N S D	N S D	*	*
	SEPT., OCT.	N S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	S D	N S D	N S D	*	*
SUSPENDED SOLIDS	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN.	N S D	N S D	N S D	*	*	N S D	N S D
	JUL, AUG	N S D	N S D	N S D	*	*	N S D	N S D
	SEPT., OCT.	S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	N S D	N S D	N S D	*	*

EXCEPT IN CASE OF D O WHERE LOW VALUES WERE USED.

T A B L E 1 A (CONTINUED)

S T A T I S T I C A L   R E S U L T S   W I T H   G R O U P I N G
(USING MAXIMUM VALUES)#

D U F F I N   C R E E K

WATER QUALITY PARAMETERS	GROUPS	RESULTS						
		PARAMETRIC			NON – PARAMETRIC			
		1-WAY	2 – WAY		FRIEDMANN	KRUSHAL-WALLIS	MANN – WHITNEY	WALD – WOLFWITZ
			DIFF. IN COL.(B's)	DIFF.IN ROW A's				
NITRATE & NITRITE	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN., JUL, AUG.	N S D	N S D	S D	N S D	N S D	*	*
	SEPT., OCT.	N S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	S D	N S D	N S D	*	*
KJELDAHL NITROGEN	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN., JUL., AUG.	N S D	N S D	N S D	N S D	N S D	*	*
	SEPT., OCT.	N S D	N S D	N S D	*	*	N S D	N S D
	NOV., DEC., JAN.	N S D	N S D	N S D	N S D	N S D	*	*
TOTAL PHOSPHATE	FEB., MAR., APR.	N S D	N S D	N S D	N S D	N S D	*	*
	MAY, JUN.	N S D	N S D	N S D	*	*	N S D	N S D
	JUL., AUG., SEPT., OCT.	N S D	N S D	N S D	N S D	N S D	*	*
	NOV., DEC., JAN.	N S D	N S D	N S D	N S D	N S D	*	*

N S D — NO SIGNIFICANT DIFFERENCE

S D — SIGNIFICANT DIFFERENCE

* — TEST NOT APPLICABLE

— EXCEPT IN CASE OF D O WHERE LOW VALUES WERE USED.

TABLE 2

ANALYSIS OF VARIANCE (1 - WAY)
(USING ALL VALUES)
$\alpha = 0.05$
DUFFIN CREEK

WATER QUALITY PARAMETER	GROUPS	NO. OF READINGS PER MONTH				SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F_A CAL.	F TAB.	REMARKS
DISSOLVED OXYGEN	FEB., MAR., APR.	8	7	7		4.22	2	2.11	0.324	3.52	N S D
	MAY, JUN.	6	6			4.94	1	4.94	2.215	4.96	N S D
	JUL., AUG., SEPT.	9	7	6		3.34	2	1.67	1.110	3.52	N S D
	OCT., NOV.	7	6			1.44	1	1.44	0.242	4.84	N S D
	DEC., JAN.	6	7			19.40	1	19.40	2.665	4.84	N S D
COLIFORMS	FEB., MAR., APR.	8	7	7		1.14×10^8	2	5.71×10^7	3.230	3.52	N S D
	MAY, JUN.	5	5			6.55×10^6	1	6.55×10^6	1.644	5.32	N S D
	JUL., AUG.	9	7			2.02×10^7	1	2.02×10^7	0.859	4.60	N S D
	SEPT., OCT.	6	7			1.34×10^8	1	1.34×10^8	1.509	4.84	N S D
	NOV., DEC., JAN.	5	6	7		7.54×10^6	2	3.77×10^6	0.589	3.68	N S D
TOTAL SOLIDS	FEB., MAR., APR.	8	7	7		3.54×10^4	2	1.77×10^4	0.898	3.52	N S D
	MAY, JUN.	6	5			15.7	1	15.7	0.008	5.12	N S D
	JUL., AUG.	9	7			6.58×10^3	1	6.58×10^3	0.489	4.60	N S D
	SEPT., OCT.	6	6			616	1	616	0.313	4.96	N S D
	NOV., DEC., JAN.	6	6	7		2.92×10^3	2	1.49×10^3	0.523	3.63	N S D
BOD (5-DAY)	FEB., MAR., APR.	8	7	7		34.3	2	34.3	0.833	3.52	N S D
	MAY, JUN., JUL, AUG.	6	5	9	7	92.8	3	30.9	0.904	3.03	N S D
	SEPT., OCT.	6	7			1.04	1	1.04	1.228	4.84	N S D
	NOV., DEC., JAN.	6	6	7		0.246	2	0.123	0.126	3.63	N S D

TABLE 2 (CONTINUED)

ANALYSIS OF VARIANCE (1 - WAY)

(USING ALL VALUES)

DUFFIN CREEK

WATER QUALITY PARAMETER	GROUP	No. OF READINGS PER MONTH				SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F_A CAL.	F TAB.	REMARKS
SUSPENDED SOLIDS	FEB., MAR., APR.	8	7	7		9.59×10^4	2	4.34×10^4	1.960	3.52	N S D
	MAY, JUN.	6	4			109.0	1	109.0	0.982	5.32	N S D
	JUL., AUG.	9	7			89.9	1	89.9	0.768	4.80	N S D
	SEPT., OCT.	6	7			697.0	1	697.0	5.616	4.84	S D
	NOV., DEC., JAN.	6	5	6		1.39×10^3	2	693.0	0.740	3.74	N S D
NITRATE & NITRITE	FEB., MAR., APR.	8	7	7		1.996	2	0.998	1.046	3.52	N S D
	MAY, JUN., JUL., AUG.	6	5	9	7	0.030	3	0.010	0.505	3.03	N S D
	SEPT., OCT.	6	7			12.2	1	12.200	1.034	4.84	N S D
	NOV., DEC., JAN.	6	6	7		0.194	2	0.0971	0.765	3.63	N S D
KJELDAHL NITROGEN	FEB., MAR., APR.	8	7	7		0.371	2	0.185	1.296	3.52	N S D
	MAY, JUN., JUL., AUG.	5	5	9	7	0.159	3	0.0531	0.534	3.05	N S D
	SEPT., OCT.	5	7			0.0734	1	0.0734	3.803	4.96	N S D
	NOV., DEC., JAN.	6	6	7		0.0534	2	0.0267	1.259	3.63	N S D
TOTAL PHOSPHATE	FEB., MAR., APR.	8	7	5		0.814	2	0.407	2.236	3.59	N S D
	MAY, JUN.	5	4			0.0172	1	0.0172	0.163	5.59	N S D
	JUL., AUG., SEPT., OCT.	9	7	6	7	39.400	3	13.100	1.084	2.99	N S D
	NOV., DEC., JAN.	5	6	6		0.072	2	0.036	1.136	3.74	N S D

T A B L E   2 A
S T A T I S T I C A L   P A R A M E T E R S   (USING MAXIMUM VALUES)
D U F F I N   C R E E K .

WATER QUALITY PARAMETERS	G R O U P S	RANGE	RATIO: HIGH/LOW	S U M OF SQUARES	DEGREE OF FREEDOM	MEAN SQUARE M S B	MEAN SQUARE M S E	F CALC.	F TAB
DISSOLVED OXYGEN	FEB., MAR., APR.	5.0, 13.8	2.76	4.14	2	2.07	3.90	0.53	3.89
	MAY, JUN.,	8.0, 13.0	1.63	8.80	1	8.80	1.69	5.21	5.32
	JUL., AUG., SEPT.	7.0, 10.6	1.52	1.50	2	0.75	1.62	0.46	4.26
	OCT., NOV.	4.0, 13.0	3.25	3.86	1	3.86	0.83	4.63	5.99
	DEC., JAN.	6.0, 13.2	2.20	15.13	1	15.13	2.56	5.90	5.32
COLIFORMS	FEB., MAR., APR.,	230, 21000	91.3	8.43×10^7	2	4.22×10^7	2.34×10^7	1.80	3.89
	MAY, JUN.	100, 7000	70.0	7.29×10^6	1	3.77×10^6	3.77×10^6	1.93	5.32
	JUL., AUG.,	170, 14000	82.3	2.79×10^7	1	2.79×10^7	2.97×10^7	0.94	5.99
	SEPT., OCT.	260, 36000	138.0	2.49×10^8	1	1.26×10^8	2.49×10^8	1.96	5.99
	NOV., DEC., JAN.	420, 8000	19.0	1.62×10^7	2	8.10×10^6	6.52×10^6	1.24	4.26
TOTAL SOLIDS	FEB., MAR., APR.	306, 962	3.14	9.89×10^4	2	4.95×10^4	2.07×10^4	2.38	3.89
	MAY, JUN.	234, 400	1.71	0.50	1	0.5	2.08×10^3	0.0002	5.32
	JUL., AUG.	274, 726	2.65	5.83×10^3	1	5832	2.40×10^4	0.24	5.99
	SEPT., OCT.	292, 398	1.36	682.67	1	683	1625	0.42	7.71
	NOV., DEC., JAN.	342, 494	1.45	1.92×10^3	2	958	3055	0.31	4.26
BOD (5 DAY)	FEB., MAR., APR.	0.9, 22.0	24.40	29.28	2	14.64	28.70	0.51	3.89
	MAY, JUN., JUL., AUG.	1.0, 32.0	32.00	132.60	2	44.20	58.50	0.76	3.49
	SEPT., OCT.	0.9, 3.2	3.56	0.03	1	0.03	0.56	0.05	5.99
	NOV., DEC., JAN.	2.1, 3.6	1.71	0.21	1	0.11	0.21	0.50	4.26
SUSPENDED SOLIDS	FEB., MAR., APR.	19, 753	39.6	1.55×10^5	2	77.49	29.41	2.94×10^4	3.89
	MAY, JUN.	10, 44	4.40	240	1	240.05	74.20	3.23	5.32
	JUL., AUG.	24, 44	1.83	91.13	1	91.13	39.96	2.28	5.99
	SEPT., OCT.	13, 50	3.85	394	1	394.33	112.28	3.51	5.99
	NOV., DEC., JAN.	16, 130	8.12	2.05×10^3	2	1.02×10^3	1.28×10^3	0.80	5.41
NITRATE & NITRITE	FEB., MAR., APR.	0.51, 5.00	12.50	1.54	2	0.77	1.54	0.59	3.89
	MAY, JUN., JUL., AUG.	0.00, 0.47	-	0.005	3	0.002	0.029	0.06	3.49
	SEPT., OCT.	0.00, 12.50	-	18.17	1	18.75	18.17	0.97	5.99
	NOV., DEC., JAN.	0.46, 1.28	2.79	0.06	2	0.03	0.10	0.27	4.26

TABLE 2A (CONTINUED)

STATISTICAL PARAMETERS (USING MAXIMUM VALUES)

DUFFIN CREEK

WATER QUALITY PARAMETERS	GROUPS	RANGE	RATIO: HIGH/LOW	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SQUARE MSB	MSE	F CALC.	F TAB.
KJELDAHL NITROGEN	FEB., MAR., APR.	0.40, 2.00	5.00	0.18	2	0.09	0.16	0.56	3.89
	MAY, JUN., JUL,, AUG.	0.00, 1.70	-	0.45	3	0.15	0.13	1.21	3.49
	SEPT., OCT.	0.36, 0.78	2.17	0.07	1	0.07	0.01	5.01	5.99
	NOV., DEC., JAN.	0.30, 0.78	2.60	0.02	2	0.12	0.03	4.51	4.26
TOTAL PHOSPHATE	FEB., MAR., APR.	0.16, 2.10	13.10	0.98	2	0.49	0.20	2.46	3.89
	MAY, JUN.	0.07, 0.90	12.9	0.02	1	0.02	0.16	0.10	7.71
	JUL., AUG., SEPT., OCT.	0.11, 19.0	173.0	65.85	3	21.95	21.83	1.00	3.49
	NOV., DEC., JAN.	0.12, 0.58	4.83	0.15	2	0.08	0.01	5.34	4.26

TABLE 3A

VARIATION BETWEEN COLUMNS (B's) - TWO WAY ANALYSIS

$\alpha = 0.05$

DUFFIN CREEK

WATER QUALITY PARAMETER	GROUPS	NO. OF READINGS PER MONTH	SUM OF SQUARES (S S B)	DEGREES OF FREEDOM	MEAN SQUARE (M S B)	SUM OF SQUARE (S S E)	F CALC. TWO-WAY	F TAB. TWO-WAY	REMARKS. DIFF.BETWEEN COLUMNS B's
DISSOLVED OXYGEN	FEB., MAR., APR.	5	4.35	2	2.17	42.13	0.475	4.46	N S D
	MAY, JUN.	5	3.84	1	3.84	1.26	11.174	7.71	S D
	JUL., AUG., SEPT.	4	2.75	2	1.37	10.99	2.835	5.14	N S D
	OCT., NOV.	4	1.13×10^{-2}	1	1.13×10^{-2}	27.60	0.003	10.10	N S D
	DEC., JAN.	5	4.50	1	4.50	156.8	0.458	7.71	N S D
COLIFORMS	FEB. MAR., APR.	5	1.05×10^{8}	2	5.25×10^{7}	1.96×10^{8}	2.197	4.46	N S D
	MAY, JUN.	5	8.45×10^{6}	1	8.45×10^{6}	1.19×10^{7}	2.841	7.71	N S D
	JUL., AUG.	4	2.79×10^{7}	1	2.79×10^{7}	3.37×10^{7}	2.482	10.10	N S D
	SEPT., OCT.	4	2.22×10^{8}	1	2.22×10^{8}	4.08×10^{8}	1.634	10.10	N S D
	NOV., DEC. JAN.	4	1.37×10^{7}	2	6.87×10^{6}	2.05×10^{7}	1.993	5.14	N S D
TOTAL SOLIDS	FEB., MAR. APR.	5	9.30×10^{4}	2	4.65×10^{4}	1.53×10^{4}	2.435	4.46	N S D
	MAY, JUN.	5	1.70	1	1.70	6.91×10^{3}	0.001	7.71	N S D
	JUL., AUG.	4	5.83×10^{3}	1	5.83×10^{3}	8.58×10^{4}	0.204	10.10	N S D
	SEPT., OCT.	3	683	1	683	5.40×10^{3}	0.253	18.50	N S D
	NOV., DEC., JAN.	4	2.90×10^{3}	2	1.45×10^{3}	1.41×10^{4}	0.617	5.14	N S D
BOD (5 - DAY)	FEB., MAR., APR.	5	44.10	2	22	172.60	1.022	4.46	N S D
	MAY, JUN., JUL., AUG.	4	131	3	43.7	493.83	0.797	3.86	N S D
	SEPT., OCT.	4	0.245	1	0.245	3.28	0.224	10.10	N S D
	NOV., DEC., JAN.	4	7.17×10^{-2}	2	3.58×10^{-2}	0.63	0.339	5.14	N S D

VARIATION BETWEEN COLUMNS (B's) — TWO WAY ANALYSIS

$\alpha = 0.05$

DUFFIN CREEK

WATER QUALITY PARAMETER	GROUPS	No. OF READINGS PER MONTH	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	SUM OF SQUARE (S S E)	F CALC. TWO-WAY	F TAB. TWO-WAY	REMARKS DIFF. BETWEEN
SUSPENDED SOLIDS	FEB., MAR., APR.	5	1.62×10^5	2	8.09×10^4	2.42×10^5	2.878	4.46	N S D
	MAY, JUN.	5	292	1	292	367	3.210	7.71	N S D
	JUL., AUG.	4	91.1	1	91.1	137	1.990	10.10	N S D
	SEPT., OCT.	4	406	1	406	349	3.487	10.10	N S D
	NOV., DEC., JAN.	3	2.05×10^3	2	1.02×10^3	2.90×10^3	1.413	6.94	N S D
NITRATE & NITRITE	FEB., MAR., APR.	5	3.14	2	1.57	11.3	1.199	4.46	N S D
	MAY, JUN., JUL., AUG.	4	9.13×10^{-3}	3	3.04×10^{-3}	0.11	0.258	3.86	N S D
	SEPT., OCT.	4	21.3	1	21.3	57	1.120	10.10	N S D
	NOV., DEC., JAN.	4	0.195	2	9.74×10^{-2}	0.31	3.094	5.14	N S D
KJELDAHL NITROGEN	FEB., MAR., APR.	5	0.343	2	0.172	1.60	0.860	4.46	N S D
	MAY, JUN., JUL., AUG.	4	0.496	3	0.165	1.23	1.211	3.86	N S D
	SEPT., OCT.	4	6.48×10^{-2}	1	6.48×10^{-2}	0.05	4.042	10.10	N S D
	NOV., DEC., JAN.	4	3.07×10^{-2}	2	1.54×10^{-2}	0.09	1.081	5.14	N S D
TOTAL PHOSPHATE	FEB., MAR., APR.	5	1.13	2	0.565	1.48	3.060	4.46	N S D
	MAY, JUN.	3	1.60×10^{-2}	1	1.60×10^{-2}	0.48	0.087	18.50	N S D
	JUL., AUG., SEPT., OCT.	4	67.5	3	22.5	197.99	1.023	3.86	N S D
	NOV., DEC., JAN.	4	7.43×10^{-2}	2	3.72×10^{-2}	0.11	2.051	5.14	N S D

VARIATION   BETWEEN   ROWS (A's) -   TWO-WAY   ANALYSIS
$\alpha = 0.05$
DUFFIN   CREEK

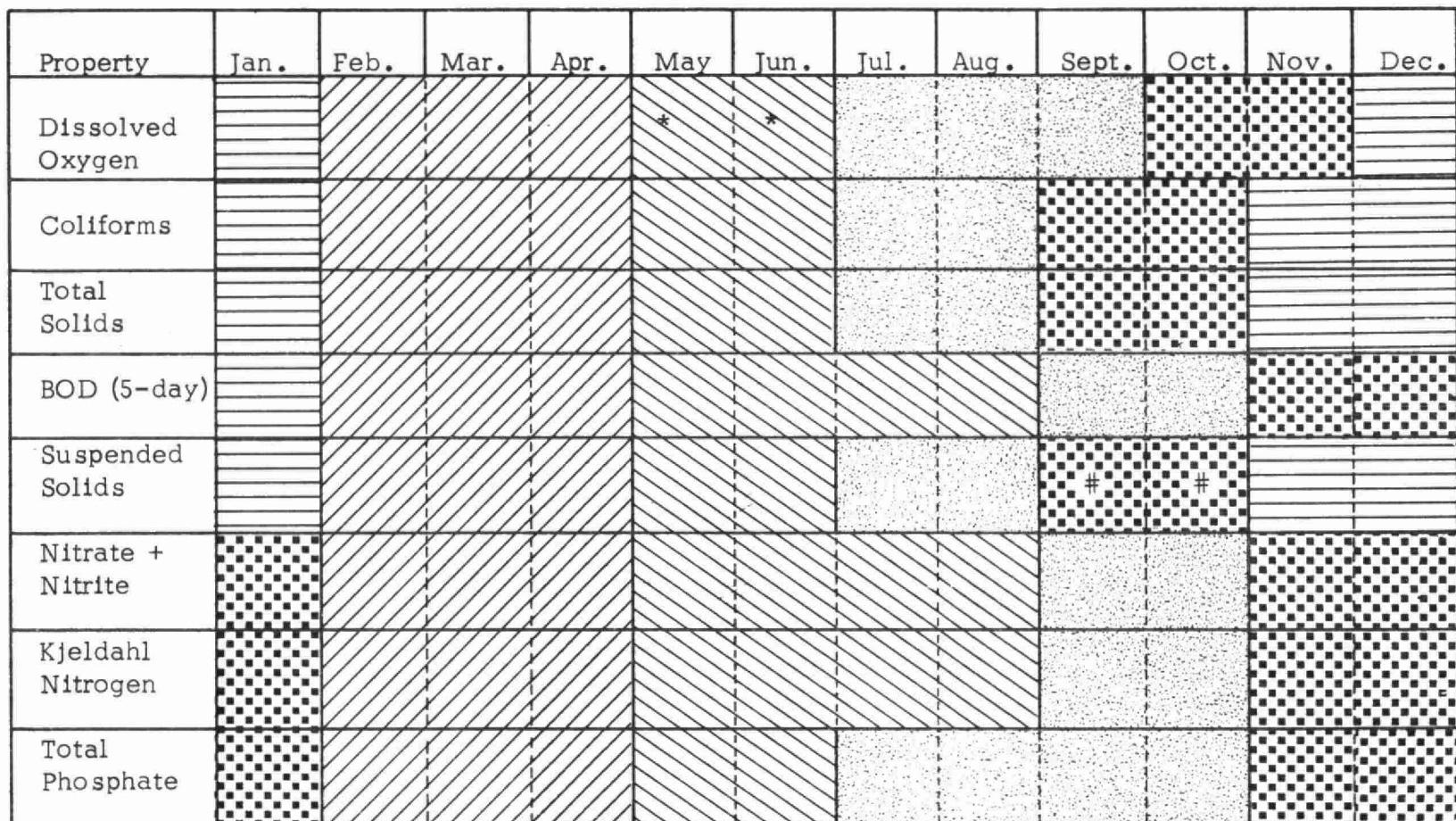| WATER QUALITY PARAMETER	GROUPS	No. OF READINGS PER GROUP	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE (M S A)	MEAN SQUARE (M S E)	F CALC. TWO-WAY	F TAB. TWO-WAY	REMARKS DIFF.BETWEEN ROWS A's
DISSOLVED OXYGEN	FEB., MAR., APR.	15	11.03	4	2.5	5.27	1.747	5.27	N S D
	MAY, JUN.	10	12.6	4	3.17	0.31	11.901	6.39	S D
	JUL,, AUG., SEPT.	12	3.7	3	1.2	1.83	3.092	4.76	N S D
	OCT., NOV.	8	37.80	3	12.6	4.35	2.899	9.28	N S D
	DEC., JAN.	8	28.10	4	9.35	9.83	0.951	6.39	N S D
COLIFORMS	FEB., MAR., APR.	15	9.10×10^7	4	2.27×10^7	2.39×10^7	0.952	3.94	N S D
	MAY, JUN.	10	1.84×10^7	4	4.59×10^6	2.97×10^6	1.543	6.39	N S D
	JUL., AUG.	8	1.45×10^8	3	4.82×10^7	1.12×10^7	4.290	9.28	N S D
	SEPT., OCT.	8	3.78×10^8	3	1.26×10^8	1.36×10^8	0.928	9.28	N S D
	NOV., DEC., JAN.	12	4.07×10^7	3	1.36×10^7	3.42×10^6	3.961	4.76	N S D
TOTAL SOLIDS	FEB., MAR., APR.	15	1.07×10^5	4	2.68×10^4	1.91×10^4	1.402	3.84	N S D
	MAY, JUN.,	10	9.74×10^3	4	2.44×10^3	1.73×10^3	1.411	6.39	N S D
	JUL., AUG.	8	5.85×10^4	3	1.95×10^4	2.86×10^4	0.682	9.28	N S D
	SEPT., OCT.	6	1.10×10^3	2	548	2.70×10^3	0.203	19.00	N S D
	NOV., DEC., JAN.	12	1.36×10^4	3	4.53×10^3	2.36×10^3	1.923	4.76	N S D
BOD (5-DAY)	FEB., MAR., APR.	15	173	4	43.2	21.57	2.005	3.84	N S D
	MAY, JUN., JUL., AUG.	16	208	3	69.3	54.87	1.263	3.86	N S D
	SEPT., OCT.	8	0.625	3	0.208	1.09	0.190	9.28	N S D
	NOV., DEC., JAN.	12	1.760	3	0.587	0.11	5.551	4.76	S D

V A R I A T I O N B E T W E E N R O W S (A's) – T W O – W A Y A N A L Y S I S

D U F F I N C R E E K

WATER QUALITY PARAMETER	GROUPS	NO. OF READINGS PER GROUP	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE (M S A)	MEAN SQUARE (M S E)	F CALC. TWO–WAY	F TAB. TWO–WAY	REMARKS DIFF. BETWEEN ROWS A's
SUSPENDED SOLIDS	FEB., MAR., APR.	15	1.21×10^5	4	3.02×10^4	3.04×10^4	1.006	3.84	N S D
	MAY, JUN.	10	247	4	61.6	90.85	0.679	6.39	N S D
	JUL., AUG.	8	102	3	34.1	45.79	0.745	9.28	N S D
	SEPT., OCT.	8	334	3	111.0	116.46	0.957	9.28	N S D
	NOV., DEC., JAN.	9	4.78×10^3	2	2.39×10^3	724.24	3.302	6.94	N S D
NITRATE & NITRITE	FEB., MAR., APR.	15	3.640	4	0.910	1.40	0.696	3.84	N S D
	MAY, JUN., JUL., AUG.	16	0.242	3	8.06×10^{-2}	0.01	6.835	3.86	S D
	SEPT., OCT.	8	55.200	3	18.5	190.00	0.974	9.28	N S D
	NOV., DEC., JAN.	12	0.504	3	0.168	0.05	5.343	4.76	S D
KJELDAHL NITROGEN	FEB., MAR., APR.	15	0.321	4	8.02×10^{-2}	0.20	0.401	3.64	N S D
	MAY, JUN., JUL., AUG.	16	0.263	3	8.78×10^{-2}	0.14	0.644	3.86	N S D
	SEPT., OCT.	8	4.11×10^{-2}	3	1.37×10^{-2}	0.02	0.854	9.28	N S D
	NOV., DEC., JAN.	12	0.158	3	5.27×10^{-2}	0.01	3.712	4.76	N S D
TOTAL PHOSPHATE	FEB., MAR. APR.	15	0.953	4	0.238	0.18	1.290	3.84	N S D
	MAY, JUN.	6	0.162	2	8.12×10^{-2}	10.24	0.341	19.00	N S D
	JUL., AUG., SEPT., OCT.	16	64.000	3	21.3	22.00	0.970	3.86	N S D
	NOV., DEC., JAN.	12	0.117	3	0.389	0.02	2.145	4.76	N S D

FINAL GROUPING

FIGURE I - DUFFIN CREEK

Property	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
Dissolved Oxygen					*	*						
Coliforms												
Total Solids												
BOD (5-day)												
Suspended Solids										#	#	
Nitrate + Nitrite												
Kjeldahl Nitrogen												
Total Phosphate												

*   Significant Difference (2-way parametric, high values)

\#   Significant Difference (1-way parametric, all values)